L1 Mosaicism in Mammals

Trends in Genetics

CellPress REVIEWS

Review

L1 Mosaicism in Mammals: Extent, Effects, and Evolution

Geoffrey J. Faulkner[1,2,*] and Jose L. Garcia-Perez[3,4]

The retrotransposon LINE-1 (long interspersed element 1, L1) is a transposable element that has extensively colonized the mammalian germline. L1 retrotransposition can also occur in somatic cells, causing genomic mosaicism, as well as in cancer. However, the extent of L1-driven mosaicism arising during ontogenesis is unclear. We discuss here recent experimental data which, at a minimum, fully substantiate L1 mosaicism in early embryonic development and neural cells, including post-mitotic neurons. We also consider the possible biological impact of somatic L1 insertions in neurons, the existence of donor L1s that are highly active ('hot') in specific spatiotemporal niches, and the evolutionary selection of donor L1s driving neuronal mosaicism.

A Mosaic of Genomes

Barbara McClintock discovered *Ac/Ds* transposition as the genetic basis for maize kernel variegation nearly 70 years ago [1,2]. In this remarkable work, McClintock simultaneously identified mobile DNA and its transposition in somatic cells, hence explaining the observed mosaic kernel phenotype. Various forms of somatic genome mosaicism have since been described [3] in normal and disease contexts, in both developing and adult tissues, involving DNA changes ranging from a single nucleotide to entire chromosomes, and in some cases these are central to crucial biological processes [4]. The mobile DNA field founded by McClintock has gone on to identify numerous transposable element (TE) families, which are arguably the preeminent feature of most eukaryotic genomes sequenced to date [5], and are a major source of genetic diversity and regulatory innovation [6,7]. However, despite an ongoing emphasis on mammalian genomics, and the instructive effects of somatic transposition on plant biology revealed by McClintock and others, our understanding of TE mobilization in mammalian somatic cells remains in its infancy. In this Review we focus on recent reports of LINE-1 (L1) retrotransposition during murine and human embryogenesis and neurogenesis, discuss the potential biological significance of somatic L1 insertions, and consider how L1 mosaicism may be subject to evolutionary selection.

L1 Retrotransposons

Retrotransposition is a molecular 'copy-and-paste' process during which an RNA template is reverse transcribed and integrated into the host genome, hence duplicating the donor DNA sequence from which the RNA was transcribed [8]. In humans, more than 500 000 L1 copies occupy ~17% of the genome [9]. An intact, full-length L1 is 6 kb in length and initiates transcription from a canonical 5′ sense promoter (Figure 1A). The L1 mRNA encodes two proteins (ORF1p and ORF2p) that catalyze L1 retrotransposition *in cis* [10]. The reverse transcriptase and endonuclease activities of ORF2p are indispensable to efficient retrotransposition [11–13]. L1 also encodes an antisense peptide, ORF0, which may assist L1 mobility [14]. Most new L1 copies are rendered immobile by 5′ truncation or internal mutation, leaving only 80–100 potentially mobile L1s per individual human genome [15,16]. Of these, fewer than 10 are expected to mobilize efficiently if tested *in vitro*, and are therefore described as 'hot' L1s

Trends

L1 retrotransposons can mobilize during embryogenesis, and in the neuronal lineage, causing somatic genome mosaicism.

Genomic analysis of endogenous L1 mobilization in mouse pedigrees, and L1 transgenic rodents, has revealed that the early embryo, before germ cell specification, is the primary niche for the accumulation of new, heritable L1 insertions.

Neuronal progenitor cells and post-mitotic neurons accommodate engineered L1 retrotransposition, but other cell lineages support limited or no L1 activity in the physiological conditions tested to date.

L1 retrotransposition clearly occurs in the brain, based on data obtained from engineered L1 reporter systems and single-cell genomic analysis, but the relevant techniques and estimated L1 mobilization rates vary considerably.

Donor L1s can be differentially active in germline and somatic cells, potentially influencing the evolutionary selection of donor L1s that are highly active in the brain.

[1]Queensland Brain Institute, University of Queensland, Brisbane, QLD 4072, Australia
[2]Mater Research Institute – University of Queensland, Translational Research Institute (TRI) Building, Woolloongabba, QLD 4102, Australia
[3]Medical Research Council Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK
[4]Department of Genomic Medicine, Centre for Genomics and Oncology (Pfizer–University of Granada and

[15,17–19]. The vast majority of hot L1s belong to one subfamily (L1-Ta, for transcribed-active) [15,17]. Although L1 is the only remaining mobile, autonomous human TE, the non-autonomous retrotransposon families *Alu* (a short interspersed element, or SINE) and SVA (a composite element incorporating SINE-R, a variable number of GC-rich tandem repeats, *Alu*, and a 5′ hexamer) can be retrotransposed *in trans* by the L1 protein machinery, as can be other polyadenylated mRNAs, generating processed pseudogenes [20–24]. In mice, ~3000 L1 copies representing three subfamilies (T_F, G_F, A) remain retrotransposition-competent (Figure 1B) as do multiple endogenous retroviruses (ERVs) and non-autonomous SINE families [25–29]. As opposed to an estimate of 1 new L1 insertion per 100 human births [30], at least 1 in 8 mice harbor a new L1 insertion [31], despite the similar percentages of identifiable L1 sequences in the mouse (~19% from 600 000 copies) and human genomes [9,32].

L1-mediated retrotransposition typically occurs via a mechanism called target-primed reverse transcription (TPRT) [33] (Box 1 and Figure 1C). As a result of TPRT, new L1 insertions typically incorporate (i) an L1 endonuclease motif, (ii) target-site duplications (TSDs), and (iii) a poly(A) tail [34]. Owing to *cis*-preference [10], L1 insertions usually arise from a retrotransposition-competent donor L1, which in some cases can be identified by L1 flanking transductions [35–37] or diagnostic internal mutations [38]. Crucially, the hallmarks of TPRT can be utilized to discriminate genuine L1 retrotransposition events from other genetic or molecular events involving L1 sequences [13,33,34,39–41]. A new L1 insertion can greatly impact on gene structure and function through insertional mutagenesis of exons [38,42,43] and regulatory elements [44], disruption of RNA polymerase II processivity [45], premature polyadenylation [46], provision of alternative promoters [14,47–51] (Figure 1A), and various other functional consequences [5,6]. L1 insertions are, likely as a result of evolutionary selection, not randomly distributed on the genome and are depleted from exons and introns [52]. This mutagenic potential also means that the L1 5′ promoter, if present in a new insertion, is heavily repressed by the host genome in most spatiotemporal contexts [48,53–56] (Figure 1D). Even in situations where full-length L1 transcripts are detected, these are usually generated by a limited number of L1 copies [38,50]. As a result, the L1 5′ promoter is a major battleground in what has often been described as an 'arms race' pitting the interests of L1 to replicate against the interest of the host genome to mitigate deleterious L1 mutations [57,58]. Beyond transcriptional repression, the host genome has developed multiple strategies to limit ongoing retrotransposition (reviewed in [59–62]).

Methods to Detect L1 Retrotransposition

Two core strategies are available to resolve the spatial and temporal extent of L1 retrotransposition: L1 reporter assays and high-throughput sequencing. In 1996, an L1 reporter assay [13] was adapted from an existing but ingenious design [8,63], and human donor L1s were

Box 1. Target-Primed Reverse Transcription (TPRT)

This seminal mechanistic model was proposed by the Eickbush laboratory [33], based on experimental data obtained from the silk moth R2 LINE-like retrotransposon, which provided a tractable system because it preferentially inserts into 28S rDNA genes [33,137]. Briefly, for murine and primate L1s, TPRT involves the transcription and translation of a full-length, capped, and polyadenylated L1 mRNA [21,139] followed by the association *in cis* of the L1 mRNA, ORF1p, and ORF2p to generate a cytoplasmic ribonucleoprotein particle (RNP) [10]. The L1 RNP can next access the nucleus [118], where the endonuclease activity of ORF2p [11] cleaves one genomic DNA strand at a degenerate 5′-TTTT/AA site [34] and then the ORF2p reverse transcriptase [12] initiates reverse transcription from the exposed 3′-hydroxyl group using the accompanying L1 mRNA as a template, generating a new L1 copy primed from the cleavage site [40]. After this process, the second DNA strand is also cleaved, presumably by the same ORF2p endonuclease activity, near to the first cleavage site which, after the nascent L1 insertion is resolved by DNA synthesis, usually leads to the formation of target-site duplications (TSDs) flanking the newly synthesized DNA. Retrotransposition can also occur through variations of the fundamental TPRT model [140–142] including occasional *trans* mobilization of mutant L1 mRNAs that do not encode intact ORFs [10,143]. The TPRT mechanism is likely to be conserved in all vertebrates because, for example, eel and zebrafish LINEs retrotranspose in human cells [144,145].

Andalusian Regional Government), Health Technology Park (PTS) Granada, 18016 Granada, Spain

*Correspondence: faulknergj@gmail.com (G.J. Faulkner).

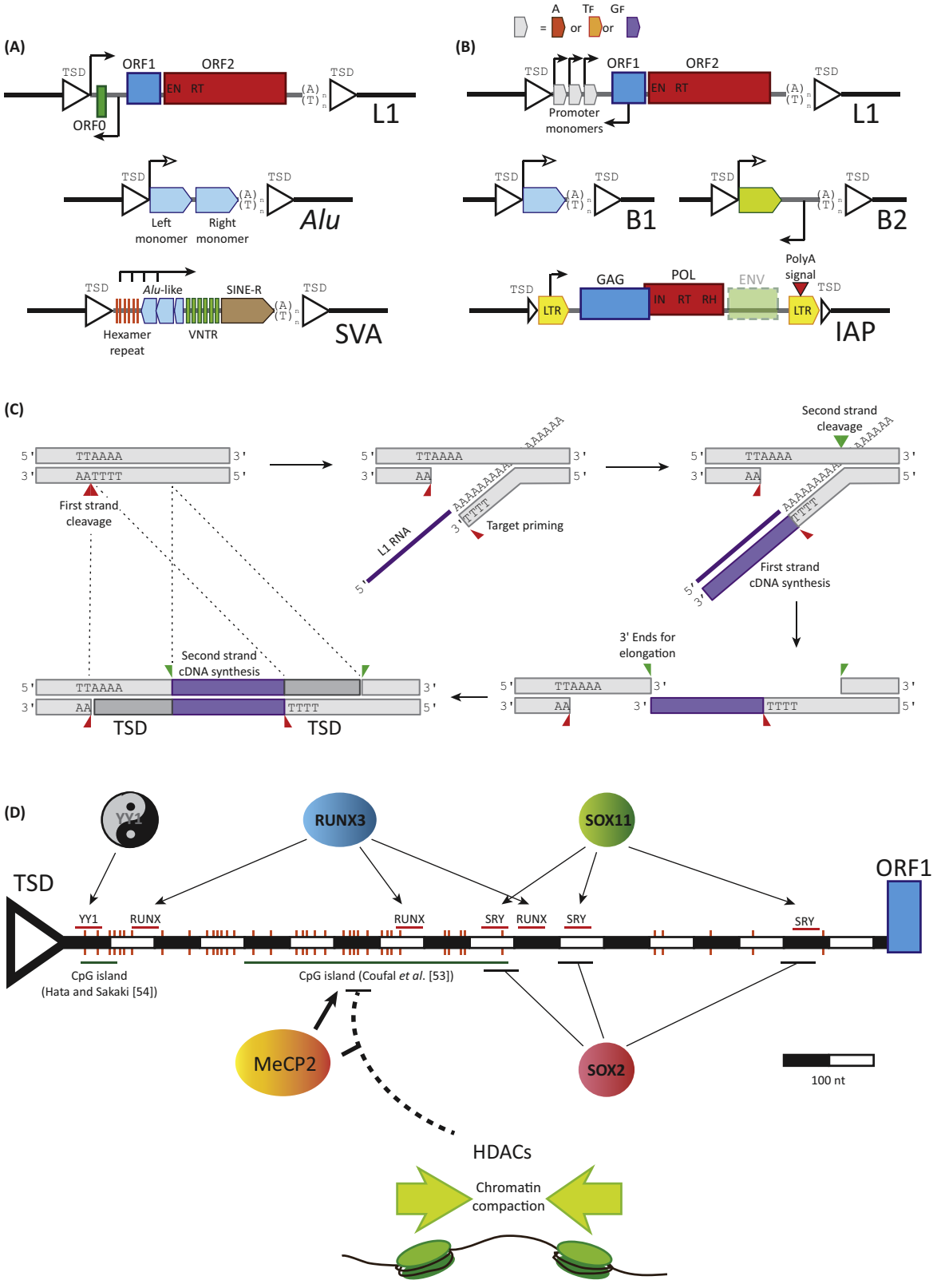

Figure 1. Mammalian Retrotransposons. (A) Mobile human retrotransposon families. L1, long interspersed element 1; *Alu*, a family of short interspersed elements (SINEs); SVA, a composite of SINE-R, variable number of tandem repeats (VNTR), *Alu*, and 5′ hexamer sequences; polymerase (Pol) II and Pol III promoters are

(See figure legend on the bottom of the next page.)

tagged with an intron-containing neomycin antibiotic resistance cassette [64] that was made functional by retrotransposition. In this assay, neomycin-resistant foci function as a readout of L1 retrotransposition efficiency (Figure 2A, left). Remarkably, frequent L1 retrotransposition events carrying TPRT hallmarks were observed in human and mouse cells (Box 1) [13]. As an alternative approach, an enhanced green fluorescent protein (EGFP)-based cassette was developed, yielding an L1–EGFP construct where EGFP was made functional by retrotransposition (Figure 2A, right) [65]. This approach facilitated the use of fluorescence-based microscopy and flow cytometry to measure L1 retrotransposition efficiency, including in transgenic animals *in vivo* [66]. In all, these reporter L1s and their derivatives (e.g., [67]) have underpinned numerous studies elucidating retrotransposon biology over the past two decades, and remain commonly used and effective tools (reviewed in [68]).

Alongside engineered L1 systems, high-throughput sequencing has massively increased our ability to characterize DNA variation in human populations [52] and cancer genomes [69]. L1 insertions are, in this regard, merely one type of DNA structural variant and can be studied *en masse*, either as part of a whole-genome sequencing (WGS) approach or via targeted sequencing of L1–genome junctions (Figure 2B). Both strategies require careful computational analysis and experimental validation to confirm true L1 insertions [39,61,70], and typically leverage L1 polymorphism catalogs [71–73] to discriminate between known and unknown L1 insertions. The bioinformatic identification of new L1 insertions from WGS data [52,74–76] is advantageous in that it can reveal the 5′ and 3′ L1–genome junctions of an insertion, allowing substantial characterization of TPRT hallmarks *a priori*. As a result, WGS analyses tend to report fewer false positives and flexibly encompass more variations of TPRT (e.g., 3′ transductions [35–37] and 5′ inversions [42,77]) than can be discerned using targeted methods [30,69,78,79] analyzing only one (usually the 3′) L1–genome junction. Some targeted methods do however attempt to analyze both L1–genome junctions simultaneously [50,80,81] and, importantly, WGS remains far more expensive than targeted approaches. Both general strategies can be applied to 'bulk' DNA extracted from tissue or pooled cells, and to DNA amplified from individual cells [82,83]. High-throughput sequencing has greatly expanded our overall capacity to study endogenous L1s *in vivo*, as opposed to the considerable caveats of introducing a transgenic L1 into a new epigenetic landscape [53,56,66,84,85]. If, however, congruent experimental data are obtained from an L1 reporter and high-throughput sequencing applied to a common biological system, such as cultured stem cells [86,87], the conclusions are likely to be robust.

Heritable and Somatic L1 Retrotransposition during Early Development

How has L1 colonized nearly one-fifth of the human and mouse genomes? Heritable L1 insertions must, by definition, occur in a germ cell or an embryonic cell contributing to the germ line. A landmark 1988 study reported L1 mutagenesis of the factor VIII gene of two hemophilia patients [42]. These results established that heritable *de novo* L1 insertions are still occurring in humans and that these mutations can cause disease. Nonetheless, the developmental origin of *de novo* L1 retrotransposition remained unclear [42]. Subsequent murine studies reported full-length L1 mRNA and L1 ORF1p expression in blastocysts, male and female germ cells, and, interestingly, in placental syncytiotrophoblast cells [88–91]. Differential

represented by solid and empty arrows, respectively. (B) As for (A), except detailing mouse L1, SINE B1, SINE B2, and IAP (intracisternal A particle) endogenous retrovirus (ERV) families. (C) Mechanism of target-primed reverse transcription (TPRT). First- and second-strand cleavage positions are depicted by red and green arrowheads, respectively. (D) Factors activating and repressing the human L1 CpG island-centric 5′-UTR promoter. CpG dinucleotides, including those assayed by two studies [53,54], are represented with vertical orange strokes. Validated transcription factor (TF) binding sites are represented by horizontal red lines [55,56,130,149,150]. Activator and repressor TFs are represented above and below the diagram, respectively. Abbreviations: EN, endonuclease; ENV, envelope; GAG, group-specific antigen; HDAC, histone deacetylase; IN, integrase; LTR, long terminal repeat; ORF, open reading frame; POL, polymerase; RH, RNase H; RT, reverse transcriptase; TSD, target-site duplication; UTR, untranslated region

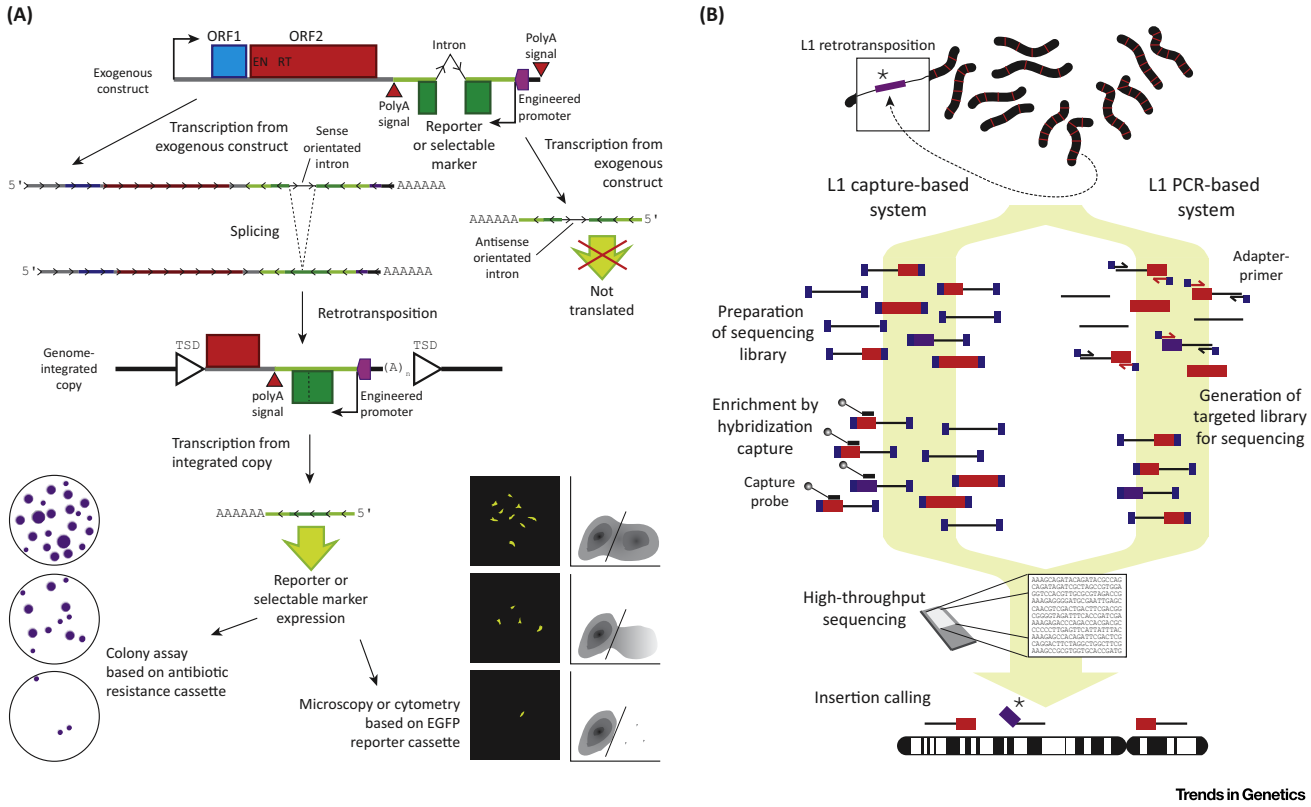

Figure 2. Methods to Identify Engineered and Endogenous L1 Insertions. (A) Schematic of an L1 reporter system. Retrotransposition from an exogenous construct carrying an L1 tagged with a spliced fluorescent reporter (e.g., EGFP [65]) or antibiotic resistance (e.g., neomycin [13]) activates the cassette, enabling downstream analysis of L1 retrotransposition efficiency. (B) Targeted sequencing approaches to map an endogenous L1 insertion. Genomic DNA can be enriched for L1–genome junctions via sequence capture [80], PCR, or adaptor ligation [30,50,79], sequenced, and computationally analyzed to reveal the *de novo* L1 variant. Abbreviations: EN, endonuclease; ORF, open reading frame; RT, reverse transcriptase; TSD, target-site duplication

L1 expression was observed during germ cell specification; for example, L1 ORF1p was detected in primordial spermatogonia, as well as in the leptotene and zygotene stages of spermatogenesis, but not in mature spermatids [88,91]. Together with later L1 transgenic mouse experiments [66,85,92–94], recovery of endogenous L1 insertions from human germ cells [95] and studies of human X-linked disease-causing L1 mutations [42,96,97], these reports strongly suggested that endogenous L1 mobilization could occur in germ cells and the early embryo.

Of highest relevance here is a study [97] that reported an L1 mutation associated with choroideremia, a rare recessive X-linked condition, in an affected male proband. Notably, his mother was a somatic and germline mosaic for the L1 insertion. This example irrefutably demonstrated that endogenous L1 retrotransposition can occur early in human embryogenesis. In addition, the *de novo* L1 insertion carried a 3′ transduction, allowing the authors to trace a full-length donor L1 and prove that it mobilized efficiently *in vitro* using the L1 reporter assay [65,97]. As a corollary, human embryonic stem cells (hESCs) support strong full-length L1 mRNA and L1 ORF1p expression [49,86,87,98–100], as do human induced pluripotent stem cells (hiPSCs) [87,100,101], human embryonic carcinoma cells [49,84], mouse embryonic stem cells (mESCs) [102], and mouse induced pluripotent stem cells (miPSCs) [101]. Consistently, the L1–EGFP reporters mobilize in hESCs, hiPSCs, and embryonic carcinoma (PA-1) cells [84,86,100], indicating that embryonic cells are likely to be a natural habitat for L1 retrotransposition.

In a recent analysis, targeted sequencing was performed on multiple cultured hESC and hiPSC lines, followed by PCR validation of candidate *de novo* insertions in multiple laboratories [87]. hiPSCs were reprogrammed from multiple parental cell types using a variety of approaches, again in several different laboratories. Eleven *de novo* L1, *Alu*, and SVA insertions were PCR-validated. These data confirmed that L1 is activated by reprogramming [100,103], a process known to involve wholesale epigenomic changes [104]. Interestingly, *de novo* L1 insertions identified in hiPSCs appeared unusually likely to be full-length, as found previously for L1–EGFP insertions in hiPSCs [100] but not in hESCs [86]. The characteristics of L1 activity may therefore be different in hiPSCs and hESCs, although an as-yet unrealized catalog of endogenous L1 insertions in cultured hESCs would be necessary to test this possibility.

In comparing the rate of endogenous L1 mobilization in hiPSCs versus hESCs, we strongly urge consideration of how heterogeneous each cell population is. Methodological factors, such as stem cell culture conditions, population bottlenecks in cultured cells, bioinformatic parameters, and how candidate L1 insertions are validated, if at all, can drastically influence results [39]. For example, a recent study [105] applied WGS to nine hiPSC lines and did not identify any *de novo* retrotransposon insertions, and far fewer mutations overall when compared to earlier studies [106,107]. Another report found 7 possible *de novo* L1 insertions in two hiPSC lines using targeted L1 sequencing, but could not PCR validate or fully characterize the genomic integration sites of these events [103]. A further study that analyzed three miPSC lines with medium coverage (10–12 $\times$ depth) WGS detected no *de novo* L1 insertions, and concluded that retroelement stability is the rule in miPSCs [108]. Given the accumulated evidence for L1 expression and mobilization in pluripotent cells, including retrotransposition of a codon-optimized L1 T_F element reporter [109–111] in mESCs (M. Garcia-Canadas *et al*., unpublished), the lack of *de novo* L1 insertions in miPSCs is perhaps surprising. There are, however, fundamental differences in how miPSCs and hiPSCs are generated and cultured, and in addition, distinct retrotransposon families appear to be more active depending on which mouse strain is analyzed [26,29,31]. Overall, we conclude that reprogramming offers L1 a dynamic but consistent relaxation of repression, and that L1 also encounters relaxed host genome control in pluripotent cells obtained directly from embryonic material [60,101,102,112]. Embryogenesis therefore provides a favorable niche for L1 retrotransposition [31].

With this in mind, WGS and targeted sequencing was recently applied to 85 mouse genomes obtained from three multigenerational C57BL/6J mouse pedigrees [31]. The developmental timing of new L1 insertions identified in progeny was then traced in parental mice, via PCR and quantitative PCR (qPCR) targeting the 5′ L1–genome junction of individual insertions. In total, 11 *de novo* insertions were identified, with all being full-length ($\geq$1 monomer) and belonging to the T_F subfamily, indicating a rate of at least one new L1 insertion per 8 births. Most heritable L1 insertions arose in the early embryo before germ cell specification or in early primordial germ cells (PGCs). For L1 insertions traced to the early embryo and early PGCs, transmission to multiple offspring was routinely observed, suggesting that more than one allele of a given event may be produced in one generation as a result of DNA replication errors and poly(A) tail shortening post-integration [113]. TE diversity within inbred strains is therefore common and adds to inter-strain variation [26,29].

Importantly, this study also identified major depletion of the 3′ L1–genome junction for the active mouse L1 families in Illumina sequencing data [31], and this was attributed to obstruction by an extensive G-quadruplex region [114,115]. To our knowledge, this issue was not identified by previous genomic analyses of mouse L1 insertions using WGS [108,116], but is potentially problematic for TE discovery and sensitivity calculations. For this reason, we consider the abovementioned figure of 1/8 to be conservative [31]. Moreover, data obtained from transgenic animals suggest that most engineered L1 retrotransposition events occur in the soma and are

not inherited [56,92,93,117]. Hence, heritable L1 insertions are likely to be far outnumbered by endogenous L1 insertions occurring in the embryo and later during ontogenesis and lineage specification.

Do Mature Neurons Support L1 Retrotransposition?

Over the past decade the L1–EGFP reporter system, alongside other approaches, has been used to elucidate engineered L1 mobilization in neural progenitors arising during fetal and adult neurogenesis (Box 2), suggesting that the brain may be a L1 mosaicism hotspot [53,56,84]. However, it remains unclear whether mature neurons, or other cell lineages, also accommodate L1 activity. Recently a human L1–EGFP reporter was introduced into hESC-derived neuronal precursor cells (NPCs) and, as seen previously [53,56], efficient retrotransposition was observed [99]. The authors then exploited a hybrid L1 adenoviral vector [118] to transduce NPCs with a modified L1–EGFP reporter, overcoming limitations associated with plasmid transfection, and again found L1 retrotransposition. Finally, the authors differentiated NPCs for 31 days to force neuronal maturation, then introduced either the adenoviral or plasmid L1–EGFP reporter along with 5-bromo-2′-deoxyuridine (BrdU), a marker of cycling cells, and found that $EGFP^+$ neuronal cells were present that were not stained with an anti-BrdU antibody. These results suggested that non-dividing neuronal cells can support extensive engineered L1 mobilization [118]. The authors performed parallel experiments in isogenic hematopoietic and mesenchymal stem cells and, compared to NPCs, observed very low L1 expression and L1–EGFP activity. Through infection with the adenoviral L1–EGFP reporter, and the use of qPCR to measure integrated EGFP copies, it was determined that the rate of L1–EGFP insertions in mature neurons was at least as high as in NPCs. These conclusions relied heavily on PCR and qPCR detection of the spliced EGFP cassette, and normalization to a plasmid or adenovirus [53]. Taken together, this study and previous studies focused on L1 in NPCs [53,56,99] lead us to conclude that engineered L1 activity, in the cell types and physiological conditions tested thus far, is largely restricted to the neuronal lineage, including post-mitotic neurons.

Extent of Endogenous L1 Mobilization in the Brain

Despite ongoing debate regarding the various types of mosaic DNA variation found in the brain [70,81,119–121], an unequivocal consensus view based on genomic analysis of bulk brain tissue [53,80], individual cells [41,81–83], and clonal cell lines derived from individual neurons [116] has formed: endogenous L1 retrotransposition can occur in the neuronal lineage, in line with foundational data obtained from engineered L1 reporter systems [53,55,56,99]. Estimates of L1 mobilization rate have nonetheless varied widely in each of the relevant studies, which have used various analytical approaches (Table 1). The earliest, and most approximate, calculations of per cell somatic L1 insertion counts were based on a L1 copy-number variation

Box 2. Engineered L1 Mobilization in Neural Progenitor Cells

A 2005 study [56] discovered *in vivo* L1–EGFP mobilization in transgenic mouse neurons, as well as in cultured rat NSCs and NPCs, providing foundational evidence of an L1 mosaic mammalian brain. Among various key findings, the authors elucidated that Sox2, a transcription factor required to maintain NSC identity [146], is a repressor of the L1 5′-untranslated region (UTR) that is downregulated to complete neuronal maturation, hence providing a scenario for L1 mobilization [56,147]. A significant caveat of this work was that it depended on a human L1 tagged with EGFP that was integrated into the foreign epigenetic landscape of another species. A subsequent study [53] showed that the L1–EGFP reporter mobilized in human NPCs derived from fetal brain and hESCs *in vitro*, and that the CpG island at the core of the L1-Ta promoter [148] was partially demethylated in fetal brain compared to non-neural tissues, further explaining L1 activation during neurogenesis. Through an L1 qPCR-based CNV assay, they determined that more L1 copies are found in brain tissues than non-brain tissues [53]. Interestingly, both studies observed that neural cells known to carry retrotransposed L1–EGFP copies could be negative for EGFP expression [53,56]. An epigenetic mechanism for transcriptional silencing of integrated L1–EGFP copies was discovered in a follow-up paper [84], suggesting that the rates of *in vitro* and *in vivo* engineered L1 retrotransposition observed in earlier work were likely to be conservative.

(CNV) assay [53], and retrotransposon capture sequencing (RC-seq) [39,80] applied to bulk hippocampal tissue, setting what appear now to be the extreme values of 80 and 0.04, respectively. In a 2012 methodological *tour de force* [82], multiple displacement amplification (MDA) was coupled to an earlier targeted 3′ L1 genome-sequencing approach [30,78] to create human-specific L1 (L1Hs) insertion profiling (L1-IP). Application of L1-IP to 300 individual pyramidal neurons from cortex and caudate nucleus revealed a single somatic L1 insertion that carried a 5′ transduction and which could be PCR amplified and capillary sequenced in its entirety (an 'empty/filled' assay, which we consider to be the highest validation standard). Another four events were detected by L1-IP but could be PCR amplified only at their 3′ L1–genome junctions. Subsequent WGS [83] applied to 16 of the MDA-amplified cortical neurons analyzed by L1-IP, including the neuron where the archetypal neuronal L1 insertion [82] was found, re-identified that event as well as another somatic L1 insertion flanked by a 614

Table 1. Summary of Endogenous L1 Mobilization Rates in Mammalian Neurons

Study	Species	Tissues	Purified neurons?	Amplification strategy[a]	Core L1 analysis method[b]	Estimated somatic L1 insertions per neuron	PCR validation details[c]	Notes	Refs
Coufal *et al.* (2009)	Human	Hippocampus, cerebellum	No	None (bulk)	L1 qPCR	80	N/A	Rate normalized to plasmid spike-in	[53]
Baillie *et al.* (2011)	Human	Hippocampus, caudate nucleus	No	None (bulk)	RC-seq	0.04	Junction-specific	Very approximate *post hoc* rate estimate [39]	[80]
Evrony *et al.* (2012)	Human	Cortex, caudate nucleus	Yes	MDA	L1-IP	0.04	Empty/filled	Validated one somatic L1 insertion carrying a 5′ transduction with empty/filled PCR, and four additional L1 insertions via junction-specific PCR	[82]
Evrony *et al.* (2015)	Human	Cortex	Yes	MDA	WGS	0.32	Empty/filled	Found somatic L1 insertion flanked by a 3′ transduction	[83]
Upton *et al.* (2015)	Human	Hippocampus, cortex	Yes	MALBAC	RC-seq	13.7	Junction-specific	Amplification method unsuitable for empty/filled PCR validation	[81]
Hazen *et al.* (2016)	Mouse	Olfactory bulb	Yes	SCNT	WGS	1.3	Junction-specific	L1 insertion sequences and families not provided	[116]
Erwin *et al.* (2016)	Human	Hippocampus, cortex	Yes	MDA	SLAV-seq	0.58–1	Empty/filled	Also identified putative somatic L1-associated deletions	[41]

[a]MALBAC, multiple annealing and looping-based amplification cycles; MDA, multiple displacement amplification; SCNT, somatic cell nuclear transfer.
[b]L1-IP, L1Hs insertion profiling; RC-seq, retrotransposon capture sequencing; SLAV-seq, somatic L1-associated variant sequencing; WGS, whole-genome sequencing.
[c]Definitions: empty/filled, PCR targeting the complete L1 insertion via amplification using primers positioned on either flank of the L1 insertion, followed by capillary sequencing (the gold standard approach); junction-specific, PCR targeting a 5′ or 3′ L1–genome junction; N/A, not applicable.

nt 3′ transduction that was, for this reason, initially overlooked by L1-IP [82]. Two additional single-cell studies of hippocampal and cortical neurons, employing MDA followed by somatic L1-associated variant sequencing (SLAV-seq) [41], and multiple annealing and looping-based amplification cycles (MALBAC) followed by RC-seq [81], also identified and PCR-validated multiple somatic L1 variants. Hence, single-cell genomic analyses have consistently found strong evidence for endogenous L1 mobilization in the neuronal lineage (Table 1).

As a discipline still in its infancy, single-cell genomic analysis can lead to conclusions heavily influenced by technical considerations. For example, a major signature of somatic C > T mutations reported by one analysis of MDA-amplified neurons [119] was suggested to be an artifact by another study employing a different genome-wide amplification approach [120]. In these circumstances, cooperation and consensus building are essential and, fortunately, the field is moving in this direction [70]. Nonetheless, the discovery and characterization of somatic L1 insertions found in a handful of cells, or even one cell, via single-cell genomics remains technically challenging [39] because whole-genome amplification and sequencing library preparation can each generate molecular artifacts, or chimeras, that obscure real L1 insertions. Sophisticated bioinformatic strategies tailored to the underlying single-cell genomic approach are hence necessary to distinguish between signal and noise. For example, variant discovery with the three targeted L1 sequencing methods used thus far to analyze neuronal genomes has filtered candidate *de novo* L1 insertions primarily based on read-count (L1-IP), L1 integration-site sequence features (single-cell RC-seq), or a combination of both read-count and sequence features (SLAV-seq) [41,81,82]. If the analysis approach suitable to one technique is applied to another (e.g., applying lessons learned from single-cell RC-seq to L1-IP [81], or the reciprocal application of a read-count filter suitable for L1-IP to single-cell RC-seq data already filtered based on sequence features [121]), the resulting L1 mobilization rate estimate can be very different, necessitating method standardization [39,70]. The common ground shared by all of these techniques is their high false positive rates, a consequent need for rigorous and time-consuming PCR validation, and their assumption that heterozygous L1 variants in single-cell genomic analysis are equivalent to somatic L1 variants [41,81,82]. This latter consideration is central to the estimation of false negatives. In this regard, it should be noted that the poly(A) tails (91 nt and 107 nt in length, on average) of the two somatic L1 insertions validated to date by the empty/filled PCR assay and presenting clear TSDs [83] are significantly longer and more adenine-pure than those carried by the vast majority of heterozygous L1 insertions [17] as a result of rapid intraindividual and intergenerational poly(A) tail shortening [83,113], and this phenomenon is even more evident for older L1 insertions [113]. Illumina sequencing is known to have issues with long homopolymer tracts [122] and it is unclear how very long poly(A) sequences fare during whole-genome amplification. Moreover, it is interesting that engineered L1 insertions have been shown to accumulate mainly in post-mitotic neurons [99], whereas the two somatic L1 insertions referred to above were each detected in multiple neurons [83]. These considerations lead us to ask whether the false negative rate has been consistently underestimated when assessing the degree of L1 mosaicism in the brain with single-cell genomics, while acknowledging that accurate false positive rate calculations are essential [81,121]. Finally, it must be noted that single-cell genomic analyses of L1 mobilization have been performed on very few human brain samples thus far, and on broad neuronal types, leaving open the possibility that some individuals, brain regions, or neuronal subtypes may support more endogenous L1 activity than others, and thus contribute to disparate somatic L1 retrotransposition frequency estimates.

To our knowledge, no single-cell analysis of endogenous L1 mobilization in the mouse brain has been reported to date. However, in an elegant study, somatic cell nuclear transfer (SCNT) was used to reprogram mESCs with neuronal nuclei obtained from the mouse olfactory bulb, followed by clonal expansion and bulk WGS to identify *de novo* TE insertions and other somatic

variants [116]. This approach provided an excellent and robust alternative to whole-genome amplification and eliminated errors associated with the latter technique, although also potentially selecting neuronal nuclei with a lower burden of DNA damage (including from L1) [116]. In six reprogrammed neuronal clones, four *de novo* L1 insertions were validated through junction-specific PCR and capillary sequencing, revealing in each case hallmark features of TPRT. Based on a false negative rate of approximately ~50%, the analyzed neurons likely each contained ~1.3 somatic L1 insertions, on average. Interestingly, this rate estimate differed dramatically from the extrema values of 0.04 [82] and 13.7 [81] obtained from single-cell genomic analysis of human neurons, although L1 appears to be more active in the mouse than in the human germline [30,31]. It is unclear how much, if at all, the 3′ L1–genome junction depletion observed recently in WGS and RC-seq data [31] affected the false negative rate calculation of this study, given that the WGS analysis appeared to group all TE families together when calculating false negative rate, and the 3′ depletion observed elsewhere was L1-specific [31]. More generally, it is unknown how much L1 activity varies in the brains of different species, or different inbred animal strains, or for that matter how much ageing and senescence impact on TE mosaicism in species with very dissimilar lifespans [123–126]. It is nonetheless remarkable that L1 mosaicism may be very common in the mouse brain, and conserved in Mammalia, based on the conservative estimate that olfactory neurons contain at least one somatic L1 insertion, on average [116].

When Does L1 Jump in Brain Development?

As noted above, engineered L1 insertions occur throughout fetal and adult neurogenesis, as well as in mature neurons [53,56,99]. With regards to endogenous L1 activity, one study detected two somatic L1 insertions, each in 2/16 neurons assayed by WGS [83]. By lineage tracing, the authors found that one of these events was timed to occur in the developing cortex and the other likely arose early in central nervous system development, and perhaps even earlier [83]. The latter circumstance reconciles well with embryonic events elucidated in mouse [31,92]. By contrast, two other studies found that most of these events appeared to arise later in neurogenesis [41,81], agreeing with reports of engineered L1 mobilization in post-mitotic neurons [99]. Interestingly, studies of engineered and endogenous L1 retrotransposition in brain tissues and neural cells have recurrently found L1 insertions in neuronal genes [41,53,56,80,81] and enhancers active in neuronal stem cells (NSCs) [81], raising the prospect of integration patterns specific to the neuronal epigenetic landscape, or post-integration selection. Pyramidal and other neuronal subtypes have been shown to contain somatic L1 insertions [81–83], as have, in far fewer instances, glia [41,81]. It therefore remains unclear whether most somatic L1 insertions found in the brain arose in the embryo, during neurogenesis, in mature post-mitotic neurons, or, as is possible, in each of these scenarios, leading to complex neuronal mosaicism.

A Model for Evolutionary Selection of Somatic L1 Retrotransposition

Is an L1 mosaic brain functionally distinct from an L1 homogenous brain? We note here only that (i) neuronal circuitry is highly interconnected and exquisitely sensitive to perturbation [127], (ii) intragenic L1 insertions can grossly impact on gene function [42,45], (iii) despite this, the potential roles of L1 mosaicism in learning and cognition remain almost entirely theoretical [39,128,129], and (iv) abnormal somatic L1 activity in neurological disorders, including Rett syndrome (RTT) [55,130], schizophrenia [131], and ataxia telangiectasia [132], has also been considered extensively but, apart from RTT, the related etiological contribution of L1 to disease is very unclear. Even for RTT, where MeCP2, a major L1 transcriptional repressor [130], is mutated and L1 mRNA, L1 protein, and L1–EGFP transgene activity are all elevated [55], MeCP2 conditional rescue can restore apparently normal neurobiological function in mice [133], meaning that L1 mosaicism is unlikely to be a major component of RTT neuronal

phenotype. Hence, although L1 insertions can impact on phenotype in the context of disease [38,42,43], it remains wholly unclear whether this applies to normal or abnormal neurobiology.

Somatic mutations are, of course, not inherited. However, donor L1s causing *de novo* L1 insertions in somatic cells are carried through the germline, and are therefore subject to selection because they can simultaneously cause germline and somatic mosaicism [31,38]. Moreover, if a particular donor L1 is very active in somatic cells it may affect the immediate evolutionary fitness of the carrier individual through disease [38,43] or even positive developmental or neurological consequences, if they exist [129]. That some L1s are apparently more mobile in somatic cells than elsewhere is supported by the identification of donor L1s that are far more active in tumors than would be expected by their activity in the germline, such as an oft-transduced donor L1 in the *TTC28* gene of numerous cancer genomes [76]. Reciprocally, some donor L1s are sufficiently active to give rise to multiple donor L1 progeny in the human germline, but have not been found to be particularly active in cancer [17,76,134]. Finally, some donor L1s are highly active in both the germline and tumors [17,38].

Donor L1s can each have multiple alleles, which can in turn mobilize at very different rates, even in the same context [18,19]. Moreover, the same donor L1 may mobilize well in one context and not another [50]. For example, the donor L1 found previously to generate a 3′ transduction-flanked neuronal L1 insertion [83] putatively mobilized during brain development but, when tested with an L1 reporter assay, did not retrotranspose in cultured osteosarcoma cells [15]. It follows that, as more active donor L1s generate longer new L1 insertions [135], they have a higher chance of generating retrotransposition-competent L1 insertions that can be easily

Key Figure

Context-Specific Donor L1 Activity

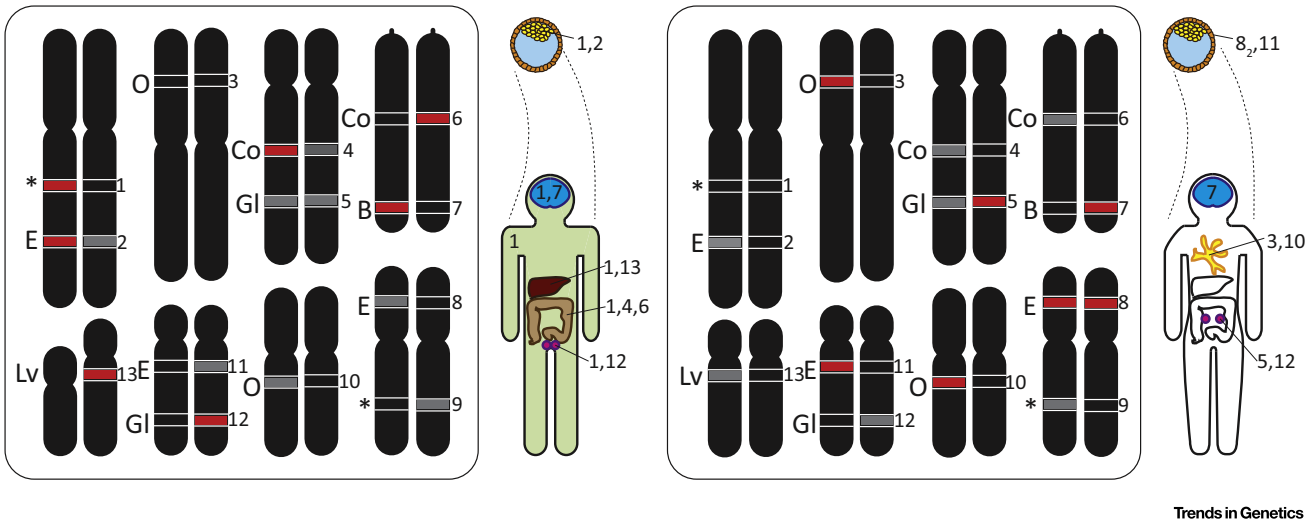

Figure 3. Schematic representation of donor L1s and their distinct impacts on germline and somatic mosaicism in two individuals. Each donor L1 locus (numbered from one to 13) can be empty (black) or contain either a retrotransposition-competent (red color) or -incompetent (grey) L1. Donor L1s can be heterozygous or homozygous. Locus-specific L1 activation can be restricted by tissue, developmental stage, or cell type (B, brain; Co, colon; E, embryo; Gl, germline; Lv, liver; O, oncological processes), or can be unrestricted (asterisk). In the anatomic panels of the respective individuals, colors and numbers represent the potential contexts of somatic L1 variants and their matched donor L1s

traced back to their donor L1s by 5′ or 3′ transduced sequences, as was the case for both of the neuronal L1 insertions referred to above [83]. For these reasons, we hypothesize that donor L1s that are 'hot' for retrotransposition [15] in particular somatic contexts *in vivo* (Figure 3, Key Figure) exist in the human population. This possibility is further supported by context-specific donor L1 activity in cultured cell lines [50], and a recent colorectal cancer study [38] that found a tumorigenic L1 insertion in the *APC* gene and traced that mutation to a polymorphic donor L1 that was demethylated not only in the tumor but also the matched normal colon [17]. If other polymorphic donor L1s are highly active in the brain, and L1 mosaicism is ultimately found to impact on neurobiology, we predict that donor L1s, the regulatory elements they carry (e.g., antisense promoters [14,48]), and the relevant host defense factors may undergo genetic selection due to their activity in the soma. Despite somatic L1 insertions not being inherited, this model could lead to varying rates of L1 mosaicism among individuals, and thus subject the phenomenon to natural selection.

Concluding Remarks and Future Directions

Endogenous L1 retrotransposition occurs in the embryo and during neurogenesis, and causes somatic genome mosaicism in neurons. The character of this mosaicism, in terms of complexity and impact, remains largely undefined. However, because the average human brain contains 80–100 billion neurons [136], even the most conservatives estimates of neuronal L1 mosaicism extrapolate to a very extensive catalogue of L1-driven variation within any individual. We would also expect that some neuronal subtypes support more L1 activity than others, perhaps as a function of when during life those neurons arise, their spatial distribution in the brain, or their neurobiological function, and in those neurons the potential for L1 insertions to drive phenotypic diversity is arguably higher than in cells that carry few or no somatic L1 insertions. TE mobilization in somatic cells is, of course, not restricted to mammals, with McClintock's maize [1,2], silk worm [137], and fruit fly [125,138] each providing examples of mosaicism caused by mobile DNA. Major advances in single-cell genomic analysis and high-throughput sequencing therefore leave the field well placed to further define somatic genome mosaicism, and its potential functional consequences, in different species and biological contexts (see Outstanding Questions).

Outstanding Questions

Human pluripotent stem cells obtained via reprogramming or from embryonic material consistently support L1 retrotransposition, as do mouse embryonic stem cells. However, it is unclear as to why *de novo* L1 insertions apparently do not occur in miPSCs. Is this a technical issue – or a result of mouse L1s being less amenable to jumping during reprogramming than human L1s (despite seeming more active in the early embryo)? Moreover, although endogenous L1 retrotransposition is now well demonstrated in the mouse embryo, it is less well defined in terms of spatial extent and frequency in early human embryogenesis.

What is the frequency of endogenous L1 mobilization in the brain? It is accepted that L1 can jump in the brain; however, the available rate estimates, and interpretations of the same data, vary widely. A focus on false positives should be complemented by a closer examination of false negatives, and standardization of techniques. L1 insertions are likely to occur in post-mitotic neurons, meaning that even a low rate of neuronal L1 mobilization could generate a constellation of L1 variation among the $\sim 10^{11}$ neurons in the human brain. Does mosaicism vary among different neuronal subtypes?

What are the immediate and broader functional consequences of somatic L1 insertions in the brain? Transcriptomic and genomic analysis of the same individual neuron could, at least, answer the first question. The impact of L1 mosaicism on neurobiology is a much more challenging and large-scale issue, with little clear evidence produced to date of somatic L1 insertions impacting on neurological function, psychiatric disorders, or neurodegenerative diseases.

If, however, L1 mosaicism impacts on neurobiology, it is plausible that donor L1s highly active in the neuronal lineage may undergo evolutionary selection despite their offspring somatic L1 insertions not being heritable. The available experimental data suggest that some donor L1s are unusually active in cancer genomes. The same may be true of donor L1s in normal somatic cells, including neurons. Hence, the donor L1 cohort of individuals, and their haplotypes, may define

Acknowledgments

G.J.F. acknowledges the support of a CSL Centenary Fellowship, National Health and Medical Research Council (NHMRC) Project Grants (GNT1106206, GNT1125645, GNT1126393) and the Mater Foundation. J.L.G-P. acknowledges funding from CICE-FEDER-P12-CTS-2256, Plan Nacional de I+D+I 2013–2016 (FIS-FEDER-PI14/02152), PCIN-2014-115-ERA-NET NEURON II, the European Research Council (ERC-Consolidator ERC-STG-2012-233764), an International Early Career Scientist grant from the Howard Hughes Medical Institute (IECS-55007420), and The Wellcome Trust–University of Edinburgh Institutional Strategic Support Fund (ISFF2). We thank Francisco Sanchez-Luque for helpful discussions and figure design, and Sandra Richardson and Adam Ewing for critical review of the manuscript.

References

1. McClintock, B. (1951) Chromosome organization and genic expression. *Cold Spring Harb. Symp. Quant. Biol.* 16, 13–47
2. McClintock, B. (1950) The origin and behavior of mutable loci in maize. *Proc. Natl. Acad. Sci. U. S. A.* 36, 344–355
3. Campbell, I.M. *et al.* (2015) Somatic mosaicism: implications for disease and transmission genetics. *Trends Genet.* 31, 382–392
4. Hozumi, N. and Tonegawa, S. (1976) Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. *Proc. Natl. Acad. Sci. U. S. A.* 73, 3628–3632
5. Levin, H.L. and Moran, J.V. (2011) Dynamic interactions between transposable elements and their hosts. *Nat. Rev. Genet.* 12, 615–627
6. Chuong, E.B. *et al.* (2017) Regulatory activities of transposable elements: from conflicts to benefits. *Nat. Rev. Genet.* 18, 71–86
7. Britten, R.J. and Davidson, E.H. (1969) Gene regulation for higher cells: a theory. *Science* 165, 349–357
8. Boeke, J.D. *et al.* (1985) Ty elements transpose through an RNA intermediate. *Cell* 40, 491–500
9. Lander, E.S. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
10. Wei, W. *et al.* (2001) Human L1 retrotransposition: cis preference versus trans complementation. *Mol. Cell. Biol.* 21, 1429–1439
11. Feng, Q. *et al.* (1996) Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* 87, 905–916
12. Mathias, S.L. *et al.* (1991) Reverse transcriptase encoded by a human transposable element. *Science* 254, 1808–1810
13. Moran, J.V. *et al.* (1996) High frequency retrotransposition in cultured mammalian cells. *Cell* 87, 917–927

14. Denli, A.M. *et al.* (2015) Primate-specific ORF0 contributes to retrotransposon-mediated diversity. *Cell* 163, 583–593
15. Brouha, B. *et al.* (2003) Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U. S. A.* 100, 5280–5285
16. Mills, R.E. *et al.* (2007) Which transposable elements are active in the human genome? *Trends Genet.* 23, 183–191
17. Beck, C.R. *et al.* (2010) LINE-1 retrotransposition activity in human genomes. *Cell* 141, 1159–1170
18. Lutz, S.M. *et al.* (2003) Allelic heterogeneity in LINE-1 retrotransposition activity. *Am. J. Hum. Genet.* 73, 1431–1437
19. Seleme, M.C. *et al.* (2006) Extensive individual variation in L1 retrotransposition capability contributes to human genetic diversity. *Proc. Natl. Acad. Sci. U. S. A.* 103, 6611–6616
20. Esnault, C. *et al.* (2000) Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* 24, 363–367
21. Doucet, A.J. *et al.* (2015) A 3′ poly(A) tract is required for LINE-1 retrotransposition. *Mol. Cell* 60, 728–741
22. Hancks, D.C. *et al.* (2011) Retrotransposition of marked SVA elements by human L1s in cultured cells. *Hum. Mol. Genet.* 20, 3386–3400
23. Dewannieux, M. *et al.* (2003) LINE-mediated retrotransposition of marked Alu sequences. *Nat. Genet.* 35, 41–48
24. Ahl, V. *et al.* (2015) Retrotransposition and crystal structure of an Alu RNP in the ribosome-stalling conformation. *Mol. Cell* 60, 715–727
25. Ostertag, E.M. and Kazazian, H.H., Jr (2001) Biology of mammalian L1 retrotransposons. *Annu. Rev. Genet.* 35, 501–538
26. Nellaker, C. *et al.* (2012) The genomic landscape shaped by selection on transposable elements across 18 mouse strains. *Genome Biol.* 13, R45
27. Sookdeo, A. *et al.* (2013) Revisiting the evolution of mouse LINE-1 in the genomic era. *Mob DNA* 4, 3
28. Goodier, J.L. *et al.* (2001) A novel active L1 retrotransposon subfamily in the mouse. *Genome Res.* 11, 1677–1685
29. Maksakova, I.A. *et al.* (2006) Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line. *PLoS Genet.* 2, e2
30. Ewing, A.D. and Kazazian, H.H., Jr (2010) High-throughput sequencing reveals extensive variation in human-specific L1 content in individual human genomes. *Genome Res.* 20, 1262–1270
31. Richardson, S.R. *et al.* (2017) Heritable L1 retrotransposition in the mouse primordial germline and early embryo. *Genome Res.* Published online May 8, 2017. http://dx.doi.org/10.1101/gr.219022.116
32. Waterston, R.H. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562
33. Luan, D.D. *et al.* (1993) Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* 72, 595–605
34. Jurka, J. (1997) Sequence patterns indicate an enzymatic involvement in integration of mammalian retroposons. *Proc. Natl. Acad. Sci. U. S. A.* 94, 1872–1877
35. Goodier, J.L. *et al.* (2000) Transduction of 3′-flanking sequences is common in L1 retrotransposition. *Hum. Mol. Genet.* 9, 653–657
36. Moran, J.V. *et al.* (1999) Exon shuffling by L1 retrotransposition. *Science* 283, 1530–1534
37. Pickeral, O.K. *et al.* (2000) Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res.* 10, 411–415
38. Scott, E.C. *et al.* (2016) A hot L1 retrotransposon evades somatic repression and initiates human colorectal cancer. *Genome Res.* 26, 745–755
39. Richardson, S.R. *et al.* (2014) L1 retrotransposons and somatic mosaicism in the brain. *Annu. Rev. Genet.* 48, 1–27
40. Monot, C. *et al.* (2013) The specificity and flexibility of L1 reverse transcription priming at imperfect T-tracts. *PLoS Genet.* 9, e1003499
41. Erwin, J.A. *et al.* (2016) L1-associated genomic regions are deleted in somatic cells of the healthy human brain. *Nat. Neurosci.* 19, 1583–1591
42. Kazazian, H.H., Jr *et al.* (1988) Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332, 164–166
43. Miki, Y. *et al.* (1992) Disruption of the APC gene by a retrotransposal insertion of L1 sequence in a colon cancer. *Cancer Res.* 52, 643–645
44. Shukla, R. *et al.* (2013) Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell* 153, 101–111
45. Han, J.S. *et al.* (2004) Transcriptional disruption by the L1 retrotransposon and implications for mammalian transcriptomes. *Nature* 429, 268–274
46. Perepelitsa-Belancio, V. and Deininger, P. (2003) RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat. Genet.* 35, 363–366
47. Wheelan, S.J. *et al.* (2005) Gene-breaking: a new paradigm for human retrotransposon-mediated gene evolution. *Genome Res.* 15, 1073–1078
48. Faulkner, G.J. *et al.* (2009) The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* 41, 563–571
49. Macia, A. *et al.* (2011) Epigenetic control of retrotransposon expression in human embryonic stem cells. *Mol. Cell. Biol.* 31, 300–316
50. Philippe, C. *et al.* (2016) Activation of individual L1 retrotransposon instances is restricted to cell-type dependent permissive loci. *Elife* 5, e13926
51. Speek, M. (2001) Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* 21, 1973–1985
52. Ewing, A.D. and Kazazian, H.H., Jr (2011) Whole-genome resequencing allows detection of many rare LINE-1 insertion alleles in humans. *Genome Res.* 21, 985–990
53. Coufal, N.G. *et al.* (2009) L1 retrotransposition in human neural progenitor cells. *Nature* 460, 1127–1131
54. Hata, K. and Sakaki, Y. (1997) Identification of critical CpG sites for repression of L1 transcription by DNA methylation. *Gene* 189, 227–234
55. Muotri, A.R. *et al.* (2010) L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446
56. Muotri, A.R. *et al.* (2005) Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910
57. Jacobs, F.M. *et al.* (2014) An evolutionary arms race between KRAB zinc-finger genes ZNF91/93 and SVA/L1 retrotransposons. *Nature* 516, 242–245
58. Imbeault, M. *et al.* (2017) KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* 543, 550–554
59. Friedli, M. and Trono, D. (2015) The developmental control of transposable elements and the evolution of higher species. *Annu. Rev. Cell Dev. Biol.* 31, 429–451
60. Gerdes, P. *et al.* (2016) Transposable elements in the mammalian embryo: pioneers surviving through stealth and service. *Genome Biol.* 17, 100
61. Goodier, J.L. (2016) Restricting retrotransposons: a review. *Mob DNA* 7, 16
62. Thompson, P.J. *et al.* (2016) Long terminal repeats: from parasitic elements to building blocks of the transcriptional regulatory repertoire. *Mol. Cell* 62, 766–776
63. Heidmann, T. *et al.* (1988) An indicator gene to demonstrate intracellular transposition of defective retroviruses. *Proc. Natl. Acad. Sci. U. S. A.* 85, 2219–2223
64. Freeman, J.D. *et al.* (1994) A modified indicator gene for selection of retrotransposition events in mammalian cells. *Biotechniques* 17, 46, 48–49, 52
65. Ostertag, E.M. *et al.* (2000) Determination of L1 retrotransposition kinetics in cultured cells. *Nucleic Acids Res.* 28, 1418–1423

the level of L1 activity in the embryo and brain.

66. Ostertag, E.M. *et al.* (2002) A mouse model of human L1 retrotransposition. *Nat. Genet.* 32, 655–660
67. Xie, Y. *et al.* (2011) Characterization of L1 retrotransposition with high-throughput dual-luciferase assays. *Nucleic Acids Res.* 39, e16
68. Rangwala, S.H. and Kazazian, H.H., Jr (2009) The L1 retrotransposition assay: a retrospective and toolkit. *Methods* 49, 219–226
69. Iskow, R.C. *et al.* (2010) Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261
70. McConnell, M.J. *et al.* (2017) Intersection of diverse neuronal genomes and neuropsychiatric disease: the Brain Somatic Mosaicism Network. *Science* 356
71. Mir, A.A. *et al.* (2015) euL1db: the European database of L1HS retrotransposon insertions in humans. *Nucleic Acids Res.* 43, D43–47
72. Sudmant, P.H. *et al.* (2015) An integrated map of structural variation in 2,504 human genomes. *Nature* 526, 75–81
73. Wang, J. *et al.* (2006) dbRIP: a highly integrated database of retrotransposon insertion polymorphisms in humans. *Hum. Mutat.* 27, 323–329
74. Helman, E. *et al.* (2014) Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res.* 24, 1053–1063
75. Lee, E. *et al.* (2012) Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971
76. Tubio, J.M. *et al.* (2014) Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* 345, 1251343
77. Ostertag, E.M. and Kazazian, H.H. (2001) Twin priming: a proposed mechanism for the creation of inversions in L1 retrotransposition. *Genome Res.* 11, 2059–2065
78. Badge, R.M. *et al.* (2003) ATLAS: a system to selectively identify human-specific L1 insertions. *Am. J. Hum. Genet.* 72, 823–838
79. Rodic, N. *et al.* (2015) Retrotransposon insertions in the clonal evolution of pancreatic ductal adenocarcinoma. *Nat. Med.* 21, 1060–1064
80. Baillie, J.K. *et al.* (2011) Somatic retrotransposition alters the genetic landscape of the human brain. *Nature* 479, 534–537
81. Upton, K.R. *et al.* (2015) Ubiquitous L1 mosaicism in hippocampal neurons. *Cell* 161, 228–239
82. Evrony, G.D. *et al.* (2012) Single-neuron sequencing analysis of L1 retrotransposition and somatic mutation in the human brain. *Cell* 151, 483–496
83. Evrony, G.D. *et al.* (2015) Cell lineage analysis in human brain using endogenous retroelements. *Neuron* 85, 49–59
84. Garcia-Perez, J.L. *et al.* (2010) Epigenetic silencing of engineered L1 retrotransposition events in human embryonic carcinoma cells. *Nature* 466, 769–773
85. Prak, E.T. *et al.* (2003) Tracking an embryonic L1 retrotransposition event. *Proc. Natl. Acad. Sci. U. S. A.* 100, 1832–1837
86. Garcia-Perez, J.L. *et al.* (2007) LINE-1 retrotransposition in human embryonic stem cells. *Hum. Mol. Genet.* 16, 1569–1577
87. Klawitter, S. *et al.* (2016) Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. *Nat. Commun.* 7, 10286
88. Branciforte, D. and Martin, S.L. (1994) Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol. Cell. Biol.* 14, 2584–2592
89. Malki, S. *et al.* (2014) A role for retrotransposon LINE-1 in fetal oocyte attrition in mice. *Dev. Cell* 29, 521–533
90. Packer, A.I. *et al.* (1993) A discrete LINE-1 transcript in mouse blastocysts. *Dev. Biol.* 157, 281–283
91. Trelogan, S.A. and Martin, S.L. (1995) Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc. Natl. Acad. Sci. U. S. A.* 92, 1520–1524
92. Kano, H. *et al.* (2009) L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* 23, 1303–1312
93. An, W. *et al.* (2006) Active retrotransposition by a synthetic L1 element in mice. *Proc. Natl. Acad. Sci. U. S. A.* 103, 18662–18667
94. Newkirk, S.J. *et al.* (2017) Intact piRNA pathway prevents L1 mobilization in male meiosis. *Proc. Natl. Acad. Sci. U. S. A.* 114, E5635–5644
95. Freeman, P. *et al.* (2011) L1 hybridization enrichment: a method for directly accessing de novo L1 insertions in the human germline. *Hum. Mutat.* 32, 978–988
96. Brouha, B. *et al.* (2002) Evidence consistent with human L1 retrotransposition in maternal meiosis I. *Am. J. Hum. Genet.* 71, 327–336
97. van den Hurk, J.A. *et al.* (2007) L1 retrotransposition can occur early in human embryonic development. *Hum. Mol. Genet.* 16, 1587–1592
98. Castro-Diaz, N. *et al.* (2014) Evolutionally dynamic L1 regulation in embryonic stem cells. *Genes. Dev.* 28, 1397–1409
99. Macia, A. *et al.* (2017) Engineered LINE-1 retrotransposition in nondividing human neurons. *Genome Res.* 27, 335–348
100. Wissing, S. *et al.* (2012) Reprogramming somatic cells into iPS cells activates LINE-1 retroelement mobility. *Hum. Mol. Genet.* 21, 208–218
101. Friedli, M. *et al.* (2014) Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. *Genome Res.* 24, 1251–1259
102. Walter, M. *et al.* (2016) An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. *Elife* 5, e11418
103. Arokium, H. *et al.* (2014) Deep sequencing reveals low incidence of endogenous LINE-1 retrotransposition in human induced pluripotent stem cells. *PLoS One* 9, e108682
104. Lister, R. *et al.* (2011) Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* 471, 68–73
105. Bhutani, K. *et al.* (2016) Whole-genome mutational burden analysis of three pluripotency induction methods. *Nat. Commun.* 7, 10536
106. Gore, A. *et al.* (2011) Somatic coding mutations in human induced pluripotent stem cells. *Nature* 471, 63–67
107. Hussein, S.M. *et al.* (2011) Copy number variation and selection during reprogramming to pluripotency. *Nature* 471, 58–62
108. Quinlan, A.R. *et al.* (2011) Genome sequencing of mouse induced pluripotent stem cells reveals retroelement stability and infrequent DNA rearrangement during reprogramming. *Cell Stem Cell* 9, 366–373
109. Goodier, J.L. *et al.* (2007) LINE-1 ORF1 protein localizes in stress granules with other RNA-binding proteins, including components of RNA interference RNA-induced silencing complex. *Mol. Cell. Biol.* 27, 6469–6483
110. Han, J.S. and Boeke, J.D. (2004) A highly active synthetic mammalian retrotransposon. *Nature* 429, 314–318
111. Naas, T.P. *et al.* (1998) An actively retrotransposing, novel subfamily of mouse L1 elements. *EMBO J.* 17, 590–597
112. Theunissen, T.W. *et al.* (2016) Molecular Criteria for Defining the Naive Human Pluripotent State. *Cell Stem Cell* 19, 502–515
113. Grandi, F.C. *et al.* (2013) LINE-1-derived poly(A) microsatellites undergo rapid shortening and create somatic and germline mosaicism in mice. *Mol. Biol. Evol.* 30, 503–512
114. Howell, R. and Usdin, K. (1997) The ability to form intrastrand tetraplexes is an evolutionarily conserved feature of the 3′ end of L1 retrotransposons. *Mol. Biol. Evol.* 14, 144–155
115. Sahakyan, A.B. *et al.* (2017) G-quadruplex structures within the 3′ UTR of LINE-1 elements stimulate retrotransposition. *Nat. Struct. Mol. Biol.* 24, 243–247
116. Hazen, J.L. *et al.* (2016) The complete genome sequences, unique mutational spectra, and developmental potency of adult neurons revealed by cloning. *Neuron* 89, 1223–1236
117. Babushok, D.V. *et al.* (2006) L1 integration in a transgenic mouse model. *Genome Res.* 16, 240–250
118. Kubo, S. *et al.* (2006) L1 retrotransposition in nondividing and primary human somatic cells. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8036–8041

119. Lodato, M.A. *et al.* (2015) Somatic mutation in single human neurons tracks developmental and transcriptional history. *Science* 350, 94–98

120. Chen, C. *et al.* (2017) Single-cell whole-genome analyses by linear amplification via transposon insertion (LIANTI). *Science* 356, 189–194

121. Evrony, G.D. *et al.* (2016) Resolving rates of mutation in the brain using single-neuron genomics. *Elife* 5, e12966

122. Quail, M.A. *et al.* (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 13, 341

123. De Cecco, M. *et al.* (2013) Genomes of replicatively senescent cells undergo global epigenetic changes leading to gene silencing and activation of transposable elements. *Aging Cell* 12, 247–256

124. Krug, L. *et al.* (2017) Retrotransposon activation contributes to neurodegeneration in a *Drosophila* TDP-43 model of ALS. *PLoS Genet.* 13, e1006635

125. Li, W. *et al.* (2013) Activation of transposable elements during aging and neuronal decline in *Drosophila*. *Nat. Neurosci.* 16, 529–531

126. Van Meter, M. *et al.* (2014) SIRT6 represses LINE1 retrotransposons by ribosylating KAP1 but this repression fails with stress and age. *Nat. Commun.* 5, 5011

127. Kobayashi, K. *et al.* (2016) Single-cell memory regulates a neural circuit for sensory behavior. *Cell Rep.* 14, 11–21

128. Muotri, A.R. *et al.* (2009) Environmental influence on L1 retrotransposons in the adult hippocampus. *Hippocampus* 19, 1002–1007

129. Singer, T. *et al.* (2010) LINE-1 retrotransposons: mediators of somatic variation in neuronal genomes? *Trends Neurosci.* 33, 345–354

130. Yu, F. *et al.* (2001) Methyl-CpG-binding protein 2 represses LINE-1 expression and retrotransposition but not Alu transcription. *Nucleic Acids Res.* 29, 4493–4501

131. Bundo, M. *et al.* (2014) Increased L1 retrotransposition in the neuronal genome in schizophrenia. *Neuron* 81, 306–313

132. Coufal, N.G. *et al.* (2011) Ataxia telangiectasia mutated (ATM) modulates long interspersed element-1 (L1) retrotransposition in human neural stem cells. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20382–20387

133. Guy, J. *et al.* (2007) Reversal of neurological defects in a mouse model of Rett syndrome. *Science* 315, 1143–1147

134. Macfarlane, C.M. *et al.* (2013) Transduction-specific ATLAS reveals a cohort of highly active L1 retrotransposons in human populations. *Hum. Mutat.* 34, 974–985

135. Farley, A.H. *et al.* (2004) More active human L1 retrotransposons produce longer insertions. *Nucleic Acids Res.* 32, 502–510

136. Herculano-Houzel, S. and Lent, R. (2005) Isotropic fractionator: a simple, rapid method for the quantification of total cell and neuron numbers in the brain. *J. Neurosci.* 25, 2518–2521

137. Eickbush, M.T. and Eickbush, T.H. (2011) Retrotransposition of R2 elements in somatic nuclei during the early development of *Drosophila*. *Mob DNA* 2, 11

138. Perrat, P.N. *et al.* (2013) Transposition-driven genomic heterogeneity in the *Drosophila* brain. *Science* 340, 91–95

139. Boeke, J.D. (2003) The unusual phylogenetic distribution of retrotransposons: a hypothesis. *Genome Res.* 13, 1975–1983

140. Gilbert, N. *et al.* (2005) Multiple fates of L1 retrotransposition intermediates in cultured human cells. *Mol. Cell. Biol.* 25, 7780–7795

141. Morrish, T.A. *et al.* (2002) DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nat. Genet.* 31, 159–165

142. Gilbert, N. *et al.* (2002) Genomic deletions created upon LINE-1 retrotransposition. *Cell* 110, 315–325

143. Garcia-Perez, J.L. *et al.* (2007) Distinct mechanisms for trans-mediated mobilization of cellular RNAs by the LINE-1 reverse transcriptase. *Genome Res.* 17, 602–611

144. Kajikawa, M. and Okada, N. (2002) LINEs mobilize SINEs in the eel through a shared 3' sequence. *Cell* 111, 433–444

145. Sugano, T. *et al.* (2006) Isolation and characterization of retrotransposition-competent LINEs from zebrafish. *Gene* 365, 74–82

146. Graham, V. *et al.* (2003) SOX2 functions to maintain neural progenitor identity. *Neuron* 39, 749–765

147. Tchenio, T. *et al.* (2000) Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res.* 28, 411–415

148. Swergold, G.D. (1990) Identification, characterization, and cell specificity of a human LINE-1 promoter. *Mol. Cell. Biol.* 10, 6718–6729

149. Athanikar, J.N. *et al.* (2004) A YY1-binding site is required for accurate human LINE-1 transcription initiation. *Nucleic Acids Res.* 32, 3846–3855

150. Yang, N. *et al.* (2003) An important role for RUNX3 in human L1 transcription and retrotransposition. *Nucleic Acids Res.* 31, 4929–4940